5. EVALUATION METHODS

Chairmen: Prof. D.I. Newble (Australia); Prof. H. Pauli (Switzerland)

Taxonomy of Essay Questions in Internal Medicine at the Final Examinations of Medical Studies

Sigurd H. Seim and Carl. W. Janssen.
A survey from the Medical Faculty of the University of Bergen, Norway.

The educational system at the Medical Faculty of the University of Bergen is a traditional one, divided into a pre-clinical study period of 2.5 years and a clinical period of 3.5 years. About 80 students pass the final examinations every year. The exam system is on the whole a conventional one, relying on a combination of oral examinations in clinical bedside settings, and on written papers answering essay questions within a limited number of hours and unassisted by any kind of reference literature.

This study was undertaken to provide insight into the kind of influence on medical students' learning strategies that could be deduced or expected from the taxonomic level of the exam questions at the written test. Students read and try to solve questions presented at previous exams during the final revision of their knowledge. Our basic presupposition is that these questions constitute forceful directives for the students' way of learning, especially since they have not been given any other written statements about study aims and objectives. Besides judging the taxonomy we have also looked into the contents and the format of the questions and, finally, have tried to form an opinion on relevance.

Our material is the collection of forty sets of exam questions in internal medicine, given in a twenty year period from 1962-1981. Each set consists of four to six different questions. Some questions have been repeated in identical fashion at intervals, but altogether there are about 180 different questions.

We applied the taxonomy introduced by Bloom (Table 1), stating six levels of competence, each higher level in the hierarchy presupposing mastery of the lower. There were some difficulties in determining taxonomic levels in a number of questions, partly because of low precision or vague formulations, but also because of ambivalence. This means presenting a pro-blem in a way leading up to one of the higher taxonomic levels, but then locking it to be answered on the lowest levels by specific demands of factual knowledge. In a general way, the two lower levels could, in our judgment, be merged together. The third level, application, would mostly be somewhat indeterminate, more or less like an appendix to the second level.

Table 1.

TAXONOMY (a.m. BLOOM)
1. Knowledge
2. Comprehension
3. Application
4. Analysis
5. Synthesis
6. Evaluation

The evaluation of the students' work is a joint task by one of their teachers and a general practitioner not belonging to the faculty staff. In our experience, students' answers in exam questions are rewarded with higher marks if they endeavour to reach a higher taxonomic level than that of the questions themselves. At least this holds true in the three lower levels. Students might even feel they have to choose between the lowest level and one of the higher in their answers, according to their interpretation of the text. According to the general assumption that clever students should give a display of an admirable wealth of factual knowledge, they safeguard themselves by recounting textbook presentations or stating theoretical principles at length, even if this is obviously outside the scope indicated by the question.

The results of our investigation are shown in Tables 2 - 5. They may be summarized as follows:

There is a total dominance of questions at the lowest taxonomic levels, the third level often being difficult to judge (Table 2). There are no questions on a higher level than the fourth analysis.

Table 2.

Taxonomic Level	OBSERVATION DATA Period 1962-71	Period 1971-81
Nr. in Level 1	89 (95.8%)	76 (85.4%)
Nr. in Level 2	?	?
Nr. in Level 3	?	2
Nr. in Level 4	4 (4.2%)	11 (12.3%)
Nr. in Level 4	0	0
Nr. in Level 6	0	0

Grouping the questions somewhat arbitrarily, according to format, we find the great majority in the group of free essay questions, the "on" questions as one might well call them (Table 3). Far fewer can be classified as belonging to one of the other four groups, which constitute different kinds of specified and limited essay presentations.

Table 3.

TYPES of EXAM QUESTIONS Grouped According to Format and Content	
Group I	: Free Essay
Group II	: How to Proceed / Solve
Group III	: Normative / Conditional
Group IV	: Exemplify / Describe
Group V	: Record, Present, Interpret

Some examples of group questions might demonstrate more precisely the kind of questions classified in the groups (Table 4). One pertinent comment would be to ask when students were taught how to proceed to clarify medical problems in general practice, an experience not included in the curriculum.

Table 4.

EXAMPLES OF GROUP QUESTIONS	
Group I	: On treatment of... : On causes, symptoms and signs of... on contraindications of... Give an account of...
Group II	: Finding glycosuria in A ... an old, overweight person, detail/discuss further steps of investigation
Group III	: How should you proceed to clarify... in general practice?
Group IV	: Give an example of a diet for... .. Describe the blood smear of...
Group V	: State the approximate calorie content of... Interpret the clinical value of the X-test How do you carry out the X-test?

The subject matter of the questions is shown (as far as main categories are concerned) in Table 5. Going into still further detail, we found it remarkable that questions regarding rheumatic conditions were represented by only one to two percent of the total. Even more remarkable was the total absence of questions concerning vascular disease of the brain and of questions about bronchial asthma, lung fibrosis and cancer. Further, questions with regard to prognostic evaluations were virtually absent. Diseases of the urogenital system represented ten percent of the total.

Table 5.

DISEASE CATEGORIES COVERED BY QUESTIONS.		
Diseases	Percentage of Questions	
	Period 1962-71	Period 1972-81
of heart/per. circ.	19.5%	22.5%
of blood	11.0%	13.5%
of lungs	6.5%	8.0%
of G.I. tract, liver pancreas	10.0%	18.0%
of endocrine system	13.0%	10.0%

Defining the high relevance of the questions as limited to those concerned with the ability to diagnose and treat acute life-threatening conditions, and conditions entailing the risk of serious and lasting health injury, we could demonstrate such relevance in a total of 22 of 110 questions belonging to group I (see Tables 3 and 4), all being in the two lower taxonomic levels, and in six questions belonging to group II. These could be assigned to the fourth level of the taxonomic scale.

Conclusion

We should not be misinterpreted as considering it not legitimate to test factual knowledge. This is clearly necessary and important. Our criticism pertains to the very great dominance of questions about factual knowledge, and we are concerned about the influence on students' learning strategies. We consider the written essay format appropriate to test factual knowledge of data and principles and of problem-solving ability in matters of high relevance as outlined above. Questions of lesser relevance should be answered with access to handbooks and/or tables. The subject matter of exam questions should be balanced and representative. Study aims and objectives should be stated separately from the written exam texts, which must have an obvious and close relationship to the former.